DOCUMENT RESUME

ED 178 600 TM 009 807

AUTHOR Lohmoller, Jan-Bernd
TITLE Estimating Parameters of Linear Structural Relation Models under Partial Least-Squares Criteria. Forschungsbericht 79.01.
INSTITUTION Hochschule der Bundeswehr, Munich (West Germany).
PUB DATE Apr 79
NOTE 45p.; Paper presented at the annual meeting of the American Educational Research Association (63rd, San Francisco, CA, April 8-12, 1979); German-language abstract also included

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Correlation; Factor Analysis; *Mathematical Models; *Statistical Analysis
IDENTIFIERS Canonical Analysis; Latent Structure Models; *Least Squares Analysis; Maximum Likelihood Estimation

ABSTRACT

A partial least squares method is described for estimating parameters of linear structural relation models. This method is an extension of Herman Wold's proposal for estimation parameters without distributional assumptions, using some algorithms worked out by Paul Horst. The method (LISPLS) determines a different number of latent variables from several sets of measured variables simultaneously and fits the relations of the latent variables either to a path model or to a correlation pattern. The name LISPLS indicates LISREL (linear structural relation model) under partial least squares criteria. Applications to three different models are described: canonical correlation analysis, the canonical redundancy model, and the interbattery factor model. These models may be generalized to analyze more blocks and different numbers of factors from each block. The model and its assumptions are described. Applications are shown to a path analysis and a multi-set factor analysis. (CTM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# FORSCHUNGSBERICHTE

ED178600

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jan-Bernd Lohmöller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Jan-Bernd Lohmöller

Estimating Parameters of Linear Structural Relation Models under Partial Least-Squares Criteria.

Forschungsbericht 79.01

IM009 807

**Hochschule der Bundeswehr München**

**Fachbereich**

**PAEDAGOGIK**

Werner-Heisenberg-Weg 39 D-8014 Neubiberg

HSBw

FORSCHUNGSBERICHT
79.01
FACHBEREICH PÄDAGOGIK
HOCHSCHULE DER BUNDESWEHR MÜNCHEN
D-8014 NEUBIBERG, WERNER-HEISENBERG-WEG 39

# PARAMETERSCHÄTZUNG FÜR LINEARSTRUKTURBEZIEHUNGSMODELLE UNTER PARTIELLEN KLEINSTQUADRATKRITERIEN.

VON
JAN-BERND LOHMÖLLER

Das Linearstrukturbeziehungsmodell (LISREL) stellt eine Verbindung von Faktorenanalyse und Pfadanalyse dar. Karl Jöreskog entwickelte für die Parameterschätzung die Maximum-Likelihood-Methode, die die Multinormalitätsannahme erfoderlich macht. Herman Wold schlug vor, die Parameter unter partiellen Kleinstquadratkriterien, also ohne Verteilungsannahmen, zu schätzen. Hier wird Wolds Methode entwickelt, unter Bezug auf Algorithmen von Paul Horst. Die Methode LISPLS (LInear Structural relation models under Partial Least Squares) bestimmt gleichzeitig mehrere latente Variablen aus mehreren Mengen gemessener Variablen, und paßt die Beziehungen der latenten Variablen entweder an ein Pfadmodell oder eine Korrelationsstruktur an.

Schlüsselwörter: Hauptkomponentenanalyse, Pfadanalyse, Kanonische Korrelation, Mehrbereichsfaktorenanalyse, LISREL, PLS-Methode.

Session
3100

# Estimating Parameters of Linear Structural Relation Models under Partial Least-Squares Criteria.

Jan-Bernd Lohmöller
Hochschule der Bundeswehr München, Fachbereich Pädagogik

Paper presented at the annual meeting of the
American Educational Research Association
April 1979, San Francisco

The Linear Structural Relation Model (LISREL) integrates features of factor analysis and path analysis. Karl Jöreskog has developed maximum-likelihood methods for parameter estimation, which require the multinormality assumtion. Herman Wold proposed to estimate parameters without distributional assumptions under partial least-squares criteria. Here Wold's method is developed, using some algorithms worked out by Paul Horst. The method LISPLS (LISREL under partial least-squares criteria) determines a different number of latent variables from several sets of measured variables simultaneously and fits the relations of the latent variables either to a path model or to a correlation pattern.

Keywords: Principal component analysis, path analysis, canonical correlation, multi-set factor analysis, LISREL, PLS-method.

# CONTENTS

1. Introduction 3

2. Wold's model 6
2.1 The linear structural relation model (LISREL) 6
2.2 Three two-factor models 7
2.3 Several factors from the same block 11

3. Model and assumptions 13
3.1 Model specification 13
3.2 Assumptions 14

4. Criteria and algorithm of parameter estimation 17
4.1 Partial least-squares criteria (PLS) 17
4.2 Iterative parameter estimation for parts of the model 18
4.3 Iterative parameter estimation for the total model 2o
4.4 Orthogonality of factors 22

5. Applications 24
5.1 A path model: Noonan & Wold (1977) 24
5.2 A multi-set factor analysis 26
5.3 Social class, environment and mental abilities 28

5 Discussion 31

Appendix 1: The PLS algorithm for several factors measured by the same block of indicator variables 33

Appendix 2: Transformation to obtain specified factor correlations equal to zero 35

Reference notes 37

References 37-38.

Author's adress:
Jan-Bernd Lohmöller
Fachbereich Pädagogik
Hochschule der Bundeswehr München
Werner Heisenberg-Weg 39
D-8o14 Neubiberg
Germany

# 1. INTRODUCTION

Karl Jöreskog has proposed three things, and all of them are called LISREL (1977, 1978):

- A model that covers various forms of hypotheses in behavioral sciences is the Linear Structural Relation (LISREL) model. It integrates confirmatory factor analysis and path analysis.
- A method to estimate the parameters of this model has ben proposed by Jöreskog. This model implies the assumption of multinormality of all variables considered in the model. The estimated parameters are Maximum Likelihood (ML) parameters.
- A program LISREL performs the estimation of parameters from an empirical covariance matrix.

The assumption of multinormality is a highly restrictive one. It is seldom checked and more seldom justified, but often accepted. This is because the estimation methods being possible under this assumption is mathematically elegant and the estimates have several desirable properties. But the algorithm requires the second derivatives of the parameters, and the programm is therefore storage- and time-consuming.

Herman Wold (1966a, 1966b, 1873, 1975, 1977, note 2, note 3) has proposed to solve Linear Structural Relation models under less restrictive assumptions. He has developed the Nonlinear Iterative Partial Least-Squares method (NIPALS or PLS). His principle for estimating the model parameters is a least squares approach, and is free from distributional assumptions.

The algorithm H. Wold uses is an extension of the vector iteration method or power method (Bock, 1975, p. 98). This is a very simple method to find a solution for homogeneous and nonhomogeneous equation systems as well as to solve symmetric eigenvalue problems, as shwon by von Mises and Pollaczek-Geiringer (1929). This method works also for assymetric

eigenvalues, e. g. principal component analysis of data matrices or canonical analysis of correlation matrices. Horst (1965) extended this method for the solution of multiple set factor analysis (MSFA), being a generalization of the canonical correlation analysis over more than two sets of variables. His approach is a deflation technique: After the iteration process has converged to the first canonical variate in each set, this variate is extracted out of its respective variables, and the iteration for a second variate starts on the deflated correlation matrix. Also Wold uses this deflation technique for the extraction of a second dimension from a block of variables.

The method which is proposed in this paper integrates some aspects of the methods of Jöreskog, Wold, and Horst. Thereby, canonical redundancy analysis (van den Wollenberg, 1977), interbattery factor analysis (Tucker, 1958), and multi-set factor analysis (Horst, 1965) are subsumed under Wold's PLS model and shown to be special cases. Based on Wold's PLS modelling, the method

- solves for several factors out of each set of variables simultaneously, thus extending the Horst algorithms,
- permits to extract different number of factors out of different sets, opposed to the MSFA which extracts an equal number of factors out of all sets,
- optimizes either a correlational structure of the factors (MSFA model) or a regression structure of the factors (LISREL model).

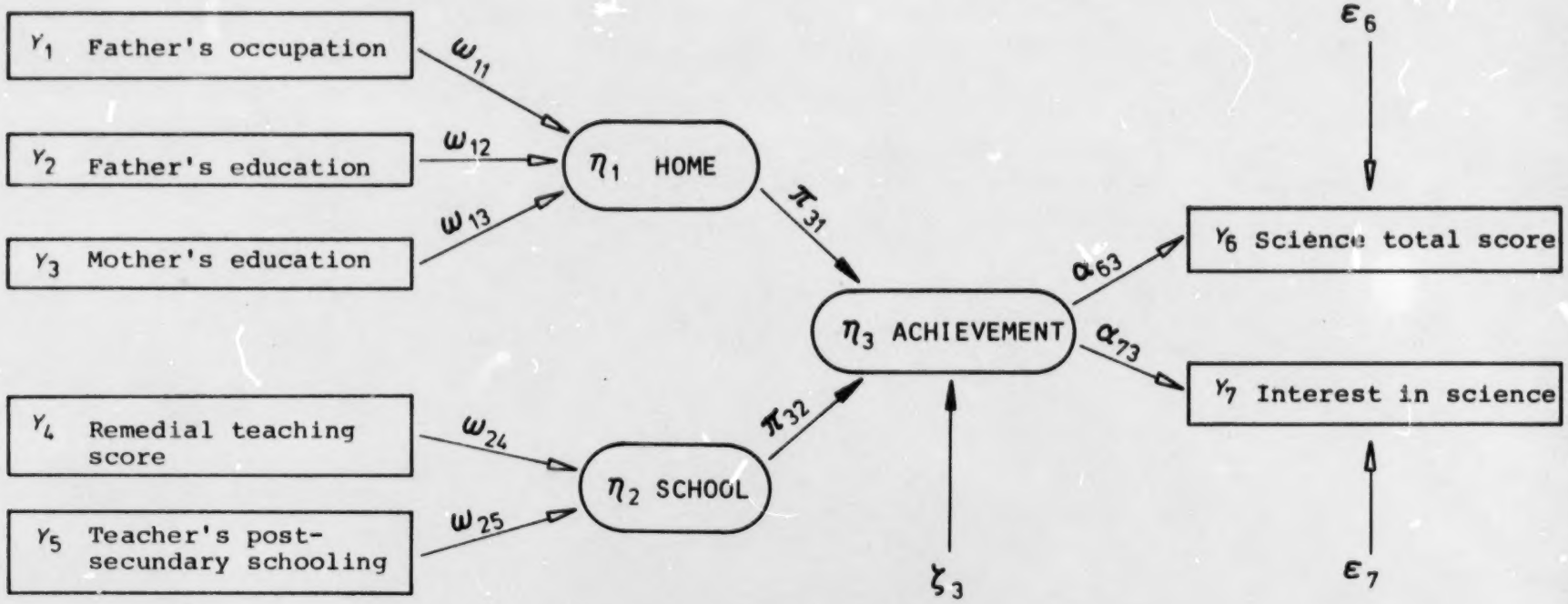


Figure 1 A linear structural relation model (from Nooan & Wold, 1977).
The block-model is $M_B = (1,1,o) \hat{=}$ (inwards,inwards,outwards)

# 2. WOLDS'S MODEL.

## 2.1 The linear structural relation model (LISREL)

Figure 1 shows the graphical representation of a linear strutural relation model. The model contains three kinds of variables:

- the structural variables $\eta$ , represented by encircled names,
- the indicator variables $Y$, represented by names in rectangles, and
- the residual variables $\zeta$ and $\varepsilon$, represented simply by names.

Two parts of the model may be distinguished, the inner part or structural model, which describes the relationships between the structural variables, and the outer parts or measurement models which describe the relationships between the structural variables and their indicators.

The inner part of the model in figure 1 contains the structural variables "home", "school" and "achievement". These are construct variables. They are not assumed to be directly measureable. They are called structural variables, or hypothetical variables, or latent variables, or even factors. The relational coefficients between the factors are labelled $\pi$ and are called path coefficients or dependency coefficients. The relationships may be interpreted by sentences like "$\eta_3$ is explained by $\eta_1$" or "$\eta_3$ is predictable from $\eta_1$" or "$\eta_3$ is influenced by $\eta_1$" or "$\eta_3$ is caused by $\eta_1$", the last two interpretations being causal ones.

The outer part of the model relate the factor $\eta_k$ to some indicator variables $Y_j$. The indicator variables are partitioned in nonoverlapping subsets or blocks. Each factor is measured by variables of only one block. The measurement model for each block may be different. There are two types of outer models, the inwards and the outwards directed measurement model. In an outwards directed block it is assumed that the factor is a

component (or predictor) of its indicator variables. It accounts for parts of the variance of its indicators. The residual variable $\varepsilon_j$ is called the measurement error. In inwards directed blocks, the factor $\eta_k$ is regressed on its indicators. There is no residual variable in this kind of measurement model. The coefficients, showing the strength of the relationships, are correlations (or, in a wider sense, factor loadings) in an outwards directed measurement model, and regression coefficients (factor weights) in an inwards directed outer model.

The model as outlined here has been proposed by Wold. The LISREL model as proposed by Jöreskog permits only outwards directed measurement models. To show the special features of Wold's model, three special models will be considered in the next section, each with two blocks of indicator variables. The models are the canonical correlation model, the canonical redundancy model, and the interbattery factor model, all of them being special cases of the Wold model. The comparison of these three two-factor models will demonstrate the substantial difference between inwards and outwards directed measurement models.

## 2.2 Three two-factor models.

The well-known model of the canonical correlation analysis assumes two structural variables $\xi$ and $\eta$, measured by two sets of variables $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ resp., and determined in such a way that the correlation $\rho = \rho(\xi, \eta)$ is maximal. The measurement models for $\xi$ and $\eta$ are inwards directed, and there is only one residual variable, called $\zeta$. The path diagramm in figure 2 can be written down as a mathematical model with three equations.

$$\xi = \sum_j \omega_j X_j \qquad (1)$$

$$\eta = \sum_k \omega_k Y_k \qquad (2)$$

$$\eta = \rho \cdot \xi + \zeta \qquad (3)$$

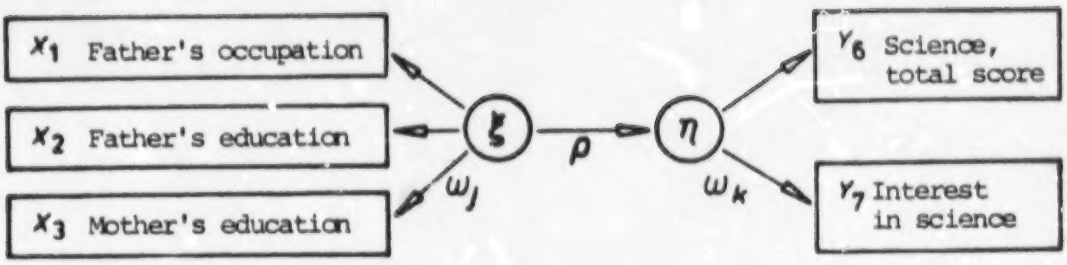


Figure 2 Canonical correlation model

Equation 1 and 2 are identities (deterministic equations, definitions). The third equation contains a residual variable. Thus, this is a stochastic equation. If $\xi$ and $\eta$ are restricted to be standardized, the correlation coefficient $\rho$ will be identical with the regression coefficient. The minimum of the variance of the residual variable $\zeta$ will be obtained if $\rho$ is maximal.

$$\rho = \max! \iff \mathrm{var}(\zeta) = \min! \qquad (4)$$

$\eta$ is the most predictable criterion (Hotelling, 1935), but we are not always interested in the most predictable criterion. Sometime we want to predict our real criterion variables. Then another two-block model applies, the canonical redundancy model (van den Wollenberg, 1977). The equations for the model are:

$$\xi = \sum_j \omega_j X_j \qquad (5)$$

$$\eta = \rho \cdot \xi \qquad (6)$$

$$Y_k = \alpha_k \cdot \eta + \varepsilon_k \qquad (7)$$

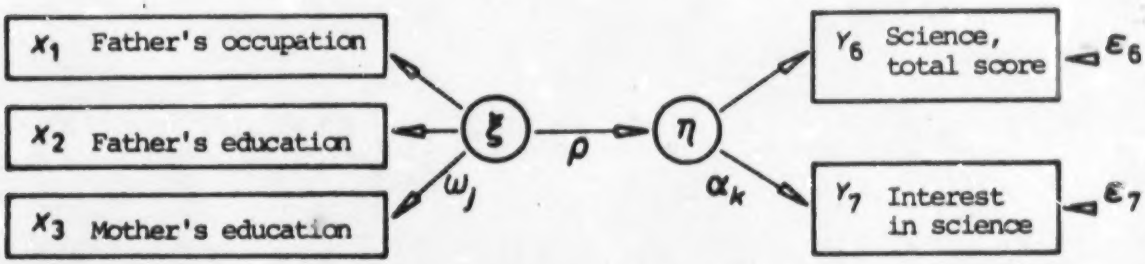


Figure 3 Canonical redundancy model.

The optimum prediction of the criterion variables $\{Y_k\}$ is guaranteed by minimizing a function $\Phi$ of the variances of all residual variables $\{\varepsilon_k\}$, say the sum of them.

$$\Phi = \Phi(\sigma^2_{\varepsilon_k}, \ldots) = \sum_k \sigma^2_{\varepsilon_k} = \text{min!} \qquad (8)$$

The scope of the canonical redundancy analysis is on the prediction. Therefore the prediction error of the real variables is minimized, and not the prediction error of a hypothetical variable as in the canonical correlation analysis.

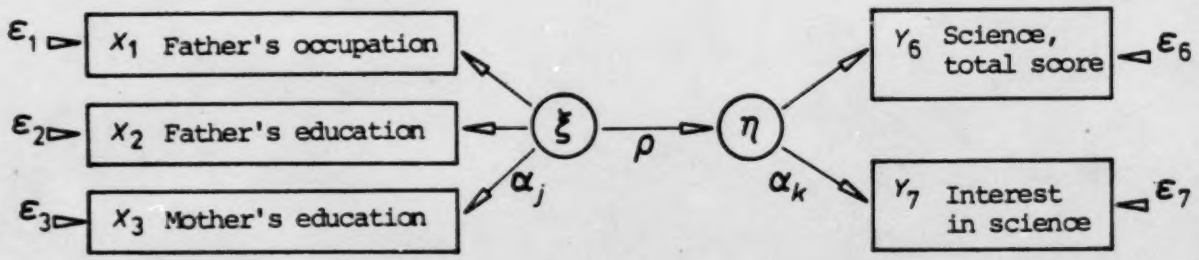


Figure 4 Interbattery factor model.

The third model to be considered is the <u>interbattery factor model</u> (Tucker, 1958), including two batteries of variables, called $\underset{\sim}{X}$ and $\underset{\sim}{Y}$, which are assumed to be influenced by the same factor. In figure 4, this common factor is split into two parts, called $\xi$ and $\eta$. The common factor is determined as to minimize the variances of all residual variables $\{\varepsilon_j\}$ and $\{\varepsilon_k\}$, or equivalently, to minimize the sum of squared loadings $\alpha_j^2$ and $\alpha_k^2$, or equivalently, the covariance of $\xi$ and $\eta$.

$$X_j = \alpha_j \cdot \xi + \varepsilon_j \qquad (9)$$

$$Y_k = \alpha_k \cdot \eta + \varepsilon_k \qquad (10)$$

$$\eta = \rho \cdot \xi \qquad (11)$$

$$\Phi = \sum_j \sigma^2_{\varepsilon_j} + \sum_k \sigma^2_{\varepsilon_k} = \text{min!} \iff \sum_j \alpha_j^2 + \sum_k \alpha_k^2 = \text{max!} \qquad (12)$$

The measurement models of the two blocks are outwards directed.

Table 1 Parameters of two-block and three-block models. Data from Noonan & Wold (1977).

| Model no. | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $M_B$ = | (1,-,1) | (1,-,o) | (o,-,o) | (1,1,1) | (1,1,o) | (o,o,o) |
| Factor loadings | | | | | | |
| Block $\eta_1$ - Home | | | | | | |
| Y1 - Father's occupation | 97 | 96 | 86 | 97 | 96 | 86 |
| Y2 - Father's education | 67 | 73 | 91 | 66 | 72 | 91 |
| Y3 - Mother's education | 75 | 8o | 91 | 75 | 8o | 91 |
| Block $\eta_2$ - School | | | | | | |
| Y4 - Remedial teaching | - | - | - | -89 | -8o | -79 |
| Y5 - Teacher's post-sec. | - | - | - | 57 | 71 | 72 |
| Block $\eta_3$ - Achievement | | | | | | |
| Y6 - Science, total score | 99 | 96 | 95 | 96 | 96 | 96 |
| Y7 - Interest in science | 44 | 79 | 8o | 39 | 78 | 79 |
| Correlations of factors | | | | | | |
| $r(\eta_1\ \eta_2)$ | - | - | - | 43 | 43 | 4o |
| $r(\eta_1\ \eta_3)$ | 74 | 66 | 62 | 74 | 67 | 63 |
| $r(\eta_2\ \eta_3)$ | - | - | - | 56 | 47 | 46 |
| Path coefficients | | | | | | |
| $p(\eta_3\ \eta_1)$ | 74 | 66 | 62 | 61 | 57 | 53 |
| $p(\eta_3\ \eta_2)$ | - | - | - | 3o | 22 | 25 |
| mult. correlation $r_3^2$ | 54 | 44 | 39 | 62 | 49 | 45 |
| variance explained (%) | 62 | 72 | 79 | 59 | 68 | 72 |

All table entries are multiplied by 1oo.
Model (1,-,1) is the canonical correlation model
Model (1,-,o) is the canonical redundancy model
Model (o,-,o) is the interbattery factor model
Model (1,1,1) is the multi-set factor model

Table 1, column 1 to 3, show that the results may be different, if data are analysed under these three different models. It can be seen that the correlation $r_{13} = {}_{est}\rho(\eta_1,\eta_3)$ of the structural variables is highest under the canonical correlation model and lowest under the interbattery factor model, whereas the proportion of explained variance (mean squared factor loadings) is lowest under the canonical correlation model and highest under the interbattery factor model.

Thus it is demonstrated that changing the direction of a measurement model can drastically influence the estimates of the parameters. Wold's model permits a choice of direction of the measurement model. From this point of view it is more flexible than Jöreskog's model. The Wold model allows to construct mathematical models more adequate to a substantial model. For this reason I am trying to work out the Wold model in some respects.

## 2.3 Several factors from the same block.

The three two-block models presented in the preceeding section may be generalized in two directions:

- Instead of two blocks of variables there may be more than two blocks.
- Instead of defining one factor from each block of indicators, one may define more than one factor from each block.

The specification of the model becomes ambiguous when more than two blocks of variables are considered. In the two-block models, there is only one relational coefficient between a pair of factors, and this coefficient may be understood to be a correlation coefficient as well as a regression coefficient. However, the relational coefficients between several factors form a matrix, and there is a difference whether this is a correlation matrix or a matrix of regression coefficients. The LISREL model designs the relationships between the factors as a path model, i.e. as a regression structure. The generalized canonical correlation (multi-set factor analysis) focusses on the correlational structure of the factors (Horst, 1965; Kettenring, 1969; Skinner, 1978).

The generalization on more than one factor per block is well-known for the two-block models and also for the multi-set factor analysis. The number of factors in these models, however, must be the same for all blocks. For the LISREL model Wold and his co-workers (Apel, 1977, in prep.; Hui, 1978) have developed a technique to determine several factors from each block of variables. After one factor per block has been estimated,

all factors are extracted from the data matrix, and then the residual matrix is used as a new input for the parameter estimation procedure. This extraction can guarantee that factors from the same block of indicators are orthogonal to each other. But nothing is known about the correlations between the first factor of the one block and another factor of another block.

This is the point from which the Wold method may be developed. The method proposed in the following pages permits to define more than one factor per block, and a different number of factors per block. This extension of the PLS method may be essential to some substantial models which include construct variables of different dimenionality, e. g. intelligence, and indicators of different reliability, e. g. age and attitudes.

Parameter estimation for the extended PLS model is done in a single run of the estimation procedure. Restrictions must be introduced to prevent that factors of the same block of indicators become identical. This can be done in three ways:

- Factors from the same block are restricted to be uncorrelated.
- Factors from the same block are bound in different path structures with factors of other blocks.
- The measurement relationships (factor weights or factor loadings) are restricted to be different, i.e. some coefficients are fixed to zero.

Up to now I have succeeded in implementing the first restriction, the restriction by zero-correlations. Before explaining how this technique works, the model must be presented in a more exact and mathematical way. This is done in the following only for the LISREL model, not for the multi-set factor model.

# 3. MODEL AND ASSUMPTIONS.

## 3.1 Model specification

Let $\underset{\sim}{Y}$ $(J\times I)$ a set of variables, where $J$ is the number of variables, and $I$ is equal to 1. If a model is specified for data, $I$ refers to the number of observations, and the expectation operator E must be replaced in the following by an average operator $1/I$ or $1/I\sum_{i=1}^{I}$ .

$\underset{\sim}{Y}$ is partitioned in nonoverlapping subsets or blocks $\underset{\sim}{Y}_l$ $(J_l\times I)$, $(l = 1,\ldots,L)$. For each block it must be specified whether it is an outwards or inwards directed block. This is done by means of a block-model matrix $\underset{\sim}{M}_B$ .

$$\underset{\sim}{M}_B = (m_l) \tag{13}$$

$$m_l = \begin{cases} 1 \text{ if block } \underset{\sim}{Y}_l \text{ is inward directed} \\ 0 \text{ if block } \underset{\sim}{Y}_l \text{ is outwards directed.} \end{cases} \tag{14}$$

Each block $\underset{\sim}{Y}_l$ is assumed to measure one or more factors $\eta_k (k = 1,\ldots,K)$. The outer-model matrix specifies, which indicator variable belongs to which factor.

$$\underset{\sim}{M}_O = (m_{jk}) \tag{15}$$

$$m_{jk} = \begin{cases} 1 \text{ if } Y_j \text{ measures } \eta_k \\ 0 \text{ else.} \end{cases} \tag{16}$$

Obviously, there are some restrictions on the specification of the outer-model matrix, such as to preserve the block model. No factor may be measured by variables of different blocks. The block of variables belonging to $\eta_k$ is $\underset{\sim}{Y}_k$ $(J_k\times I)$.

The model equations for the measuring models are different for factors from inwards and outwards directed blocks. The model equation for an inwards directed block describes a factor as a linear composite of its indicators,

$$\eta_k = \underset{\sim}{\Omega}_k' \underset{\sim}{Y} , \tag{17}$$

where $\underset{\sim}{\Omega}_k$ is a $(1\times J)$ vector of "factor weights", and the prime

denotes the matrix transposition. The measuring equation for an outwards directed block describes the indidicator variable as a linear composite of its factor and its residual variable,

$$y_j = \alpha_{jk} \cdot \eta_k + \varepsilon_j \ , \tag{18}$$

where $\underset{\sim}{A}_k = (\alpha_{jk})$ is a $(J \times 1)$ vector of factor patterns.

The relationships between the factors are specified by the inner-model matrix $\underset{\sim}{M}_I$.

$$\underset{\sim}{M}_I = (m_{kk'}) \tag{19}$$

$$m_{kk'} = \begin{cases} 1 & \text{if } \eta_{k'} \text{ is source variable for } \eta_k \\ 0 & \text{else.} \end{cases} \tag{20}$$

A factor $\eta_k$ is called exogenous, if it has no source variable, i.e. $m_{kk'} = 0$ for $k' = 1,\ldots,K$ , or else $\eta_k$ is called endogenous. The model equation for the inner variables is

$$\underset{\sim}{\eta} = \underset{\sim}{\Pi}\underset{\sim}{\eta} + \underset{\sim}{\zeta} \tag{21}$$

$$= (\underset{\sim}{I} - \underset{\sim}{\Pi})^{-1} \underset{\sim}{\zeta} \tag{21a}$$

$$= (\textstyle\sum_{m=0}^{\infty} \underset{\sim}{\Pi}^m) \underset{\sim}{\zeta} \tag{21b}$$

$$= \underset{\sim}{\Pi}^{*} \underset{\sim}{\zeta} \ , \tag{21c}$$

where $\underset{\sim}{\Pi}$ is the $(K \times K)$ matrix of path coefficients, $\underset{\sim}{I}$ is the $(K \times K)$ identity matrix, and $\underset{\sim}{\zeta}$ is a $(K \times I)$ vector of residual variables, called the specification error. Equation 21c shows the reduced form of the model; the variables $\underset{\sim}{\eta}$ are reduced on the predetermined variables $\underset{\sim}{\zeta}$ and are not regressed on each other. The term "nonlinear PLS" refers to equation 21b, showing the factors to be nonlinear in the predetermined variables $\underset{\sim}{\zeta}$.

## 3.2 Assumptions

To simplify the treatment of all indicator variables, pattern vectors are also defined for inwards blocks, and weight vectors for the outwards blocks. They are collected in a weight matrix $\underset{\sim}{\Omega}$ $(J \times K)$ and a pattern matrix $\underset{\sim}{A} = (\alpha_{jk})$ $(J \times K)$, resp. It must be pointed out that not all elements of $\underset{\sim}{\Omega}$ and $\underset{\sim}{A}$ are parameters of the model, but only those defined by equations

17 and 18. With these matrices we have the relationships

$$\underset{\sim}{\eta} = \underset{\sim}{\Omega}' \underset{\sim}{Y} \tag{17a}$$

$$\underset{\sim}{Y} = \underset{\sim}{A}' \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \tag{18a}$$

$$= \underset{\sim}{A}' \underset{\sim}{\Pi}^{\bullet} \underset{\sim}{\zeta} + \underset{\sim}{\varepsilon} .$$

From the definition of the model matrices (equations 16 and 2o) it follows that the elementwise product of the factor weight with the outer-model matrix does not change the weight matrix, the same being true also for the pattern matrix,

$$\underset{\sim}{\Omega} = \underset{\sim}{\Omega} * \underset{\sim}{M}_O \tag{22}$$

$$\underset{\sim}{A} = \underset{\sim}{A} * \underset{\sim}{M}_O , \tag{23}$$

where the asterisk denotes the Hadamard (elementwise) product. A similar formula holds for the inner model.

$$\underset{\sim}{\Pi} = \underset{\sim}{\Pi} * \underset{\sim}{M}_I \tag{24}$$

In simple terms, the zero-elements in the model matrices correspond to zero-elements in $\underset{\sim}{\Omega}$, $\underset{\sim}{A}$ and $\underset{\sim}{\Pi}$, resp.

The assumptions on the means and the variances of the variables are

$$E(Y_j) = E(\varepsilon_j) = E(\eta_k) = E(\zeta_k) = o \tag{25}$$

$$E(\eta_k^2) = 1 . \tag{26}$$

The assumptions on the covariances of the residual variables are

$$E(\varepsilon_j \varepsilon_{j'}) = o \quad \text{if } Y_j \text{ or } Y_{j'} \text{ belongs to an outward directed block} \tag{27}$$

$$E(\zeta_k \zeta_{k'}) = o \quad \text{if } m_{kk'} \text{ or } m_{k'k} \text{ is } \neq o \tag{28}$$

$$E(\varepsilon_j \zeta_k') = o$$

We define the covariance matrices of the variables:

$$\underset{\sim}{\Theta} = E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') \tag{29}$$

$$\underset{\sim}{\Psi} = E(\underset{\sim}{\zeta}\underset{\sim}{\zeta}') \tag{30}$$

$$\underset{\sim}{P} = E(\underset{\sim}{\eta}\underset{\sim}{\eta}') = (\rho_{kk'}) \tag{31}$$

$$= E(\underset{\sim}{\Pi}^{*}\ \underset{\sim}{\zeta}\underset{\sim}{\zeta}'\ \underset{\sim}{\Pi}^{*}') \qquad \text{by eq. 21c}$$

$$= \underset{\sim}{\Pi}^{*}\ \underset{\sim}{\Psi}\ \underset{\sim}{\Pi}^{*}'$$

$$\underset{\sim}{\Sigma} = E(\underset{\sim}{Y}\underset{\sim}{Y}') \tag{32}$$

$$= E[(\underset{\sim}{A}\ \underset{\sim}{\Pi}^{*}\ \underset{\sim}{\zeta} + \underset{\sim}{\varepsilon})(\underset{\sim}{\zeta}'\ \underset{\sim}{\Pi}^{*}'\ \underset{\sim}{A}' + \underset{\sim}{\varepsilon}')] \qquad \text{by eq. 18a}$$

$$= \underset{\sim}{A}\ \underset{\sim}{\Pi}^{*}\ \underset{\sim}{\Psi}\ \underset{\sim}{\Pi}^{*}'\ \underset{\sim}{A}' + \underset{\sim}{\Theta} \tag{32a}$$

$$\underset{\sim}{B} = (\beta_{jk}) = E(\underset{\sim}{Y}\underset{\sim}{\eta}') \tag{33}$$

$$= E(\underset{\sim}{Y}\ \underset{\sim}{Y}'\ \underset{\sim}{\Omega}) \qquad \text{by eq. 17a}$$

$$= \underset{\sim}{\Sigma}\ \underset{\sim}{\Omega}$$

Equation 32a expresses the covariances of the indicator variables in outwards directed blocks in terms of the model parameters. Equation 33 defines the factor structure matrix $\underset{\sim}{B}$, i.e. the matrix of correlations of all variables with all factors. If factors from the same block of variables are uncorrelated, then

$$\underset{\sim}{A} = \underset{\sim}{B} * \underset{\sim}{M}_{O} . \tag{34}$$

# 4. CRITERIA AND ALGORITHM OF PARAMETER ESTIMATION

When estimating the parameters of a model from empirical data, the model parameters are replaced by their estimates.

$\hat{\eta} = \text{est}^{\eta}$

$\underset{\sim}{W} = \text{est}^{\underset{\sim}{\Omega}}$

$\underset{\sim}{B} = \text{est}^{(\beta_{jk})}$

$\underset{\sim}{A} = \text{est}^{(\alpha_{jk})}$

$\underset{\sim}{P} = \text{est}^{\Pi}$

$\underset{\sim}{R} = \text{est}^{(\rho_{kk'})}$

$\hat{\underset{\sim}{\Theta}} = \text{est}^{\underset{\sim}{\Theta}}$

$\hat{\underset{\sim}{\Psi}} = \text{est}^{\underset{\sim}{\Psi}}$

$\hat{\underset{\sim}{\Sigma}} = \text{est}^{\underset{\sim}{\Sigma}}$

## 4.1 Partial least-squares criteria (PLS)

The next step when specifying the model would be to set up a scalar criterion, to be optimized with respect to the parameters of the model under the restrictions imposed on them. A lot of well-known models, e.g. the principal component analysis, the multiple regression, and the canonical correlation, use a criterion related to the variance of a residual variable. However, in our model there are different residual variables, the inner residuals $\underset{\sim}{\zeta}$ and the outer residuals $\underset{\sim}{\varepsilon}$. Combining their variances in a single criterion has some arbitrariness. Summing up all residual variances would mean that inner and outer residuals are of equal importance.

Herman Wold proposed to treat each residual variance separately by defining a set of partial least-squares criteria (or local least-squares criteria):

$$\sum_{j_k} \hat{\theta}^2_{j_k} = \text{min!} \qquad \text{for all outwards directed blocks} \qquad (35)$$

$$\hat{\psi}^2_k = \text{min!} \qquad \text{for all endogenous factors} \qquad (36)$$

The first criterion says that the sum of the outer residual variances is minimized in all outwards directed blocks separately, whereas in inwards directed blocks this criterion does not apply. The second criterion says that the multiple correlation of each inner variable is maximized. The partial least squares criteria are in some special cases, e.g. principal component analysis and canonical correlation analysis, identical to other well-defined scalar criteria. In other cases, the algorithm of parameter estimation works so as to provide a compromise between all local criteria.

## 4.2 Iterative parameter estimation for parts of the model

From the fact that no single criterion has been established, it follows that no algebraic solution for the parameter estimation is possible, except for models with one or two structural variables (Wold, 1975, p. 332). So an iterative algorithm to apply on the local criteria must be found.

The first basic idea of the partial least-squares parameter estimation technique is that one part of the parameters is considered as known and held fixed, while the other part of the parameters is estimated. There are only so many parameters included in the one part of parameters to be estimated, that a least-squares solution is possible. The term "partial least squares" refers to the partitioning of the parameters in estimable subsets. Three typical parts of a model will be considered which permit a local solution.

The multiple regression. If $Z$ is a target variable (criterion variable) and $X_j$ $(j = 1,\ldots,J)$ is a set of source (predictor) variables, then the least-squares approximation to $Z$ is known to be

$$_{c+1}Z = \sum_j w_j X_j \ , \qquad (37)$$

where the weights $w_j$ are the regression coefficients.

The principal component. If $Z$ is the principal component of a set of target variables $Y_k$ $(k = 1,\ldots,K)$ and $_cZ$ is an approximation to $Z$, then this approximation is improved by the linear combination

$$_{c+1}Z = {}_cf \sum_k {}_ca_k Y_k \ , \qquad (38)$$

where $_ca_k$ is the correlation coefficient $r(_cZ, Y_k)$ and $_cf$ is a standardization factor. To obtain the principal component, the computation (38) must be repeated for iteration cycle $c = 1,\ldots$ until the difference $_{c+1}a_k - {}_ca_k$ $(k = 1,\ldots,K)$ is judged to be neglegible (cf. Bock, 1975, p. 93).

The MIMIC model. Consider a model with multiple indicators (effect variables, target variables) $Y_k$ $(k = 1,\ldots,K)$ and multiple causes (source variables) $X_j$ $(j = 1,\ldots,J)$ of a single latent variable $Z$ (MIMIC model, cf. Jöreskog & Goldberger, 1975). If $_cZ$ is an approximation to $Z$, then this approximation is improved by the linear combination

$$_{c+1}Z = {}_cf(\sum_j {}_cw_j X_j + \sum_k {}_ca_k Y_k) \ , \qquad (39)$$

where $_cw_j$ is a multiple regression coefficient $w(_cZ, X_j)$, $_ca_k$ is the correlation coefficient $r(_cZ, Y_k)$, and $_cf$ is a standardization factor. To obtain the least-squares estimator of $Z$, the computation (39) must be repeated for iteration cycle $c = 1,\ldots$ until the differences between successive parameter estimates are judged to be neglegible (cf. Wold, 1975, p. 327).

Summing up, it can be said that the linear composite $Z$ will tend to be

- the most predictable criterion, if the regression coefficients are used as weights (equation 37),
- the principal component or best predictor, if the corre-

lation coefficients are used as weights (equation 38),
- the best explanatory variable, if the regression coefficients are used as weights of the source variables, and the correlation coefficients as weights of the target variables (equation 39).

Estimation problems as considered occur as local problems in a PLS model. The multiple regression problem occurs when estimating the factor of an inward directed block, the principal component problem when estimating the factor of an outward directed block, and all three problems occur in the inner part of the model.

## 4.3 Iterative parameter estimation for the total model

The second basic idea of the PLS procedure is that the parameters of a partial model are not solved totally, but that each local estimate is improved one step in each iteration cycle. The PLS algorithm can be extended in order to determine several uncorrelated factors from the same block of variables. To show the way, how the algorithm can be extended, the basic algorithm will be presented first (cf. appendix 1).

Each factor $\eta_k$ is involved in two equation systems, the system of "definition equation" (eq. 17a, 18a) which relate it to the indicator variables $Y_j$, and the "structural equation" (eq. 21) which relate it to the other factors. The PLS algorithm uses both equation system in turn. First an "outside approximation" $_o\tilde{\eta}_k$ to the factor $\eta_k$ is computed by using the definition equation, thus defining the factor in terms of its indicator variables. The weights are either regression coefficients, if $\eta_k$ is measured by an inwards directed block, or correlation coefficients, if it is measured by an outwards directed block of indicator variables. In the second step an "inside approximation" $_i\tilde{\eta}_k$ is computed using the structural equations, thus defining the factor $\eta_k$ in terms of the other structural variables.

In the third step new weights for the measurement relations are computed. This is done by regressing the inner estimates ${}_o\ddot{\eta}_k$ on the indicator variables. In an inwards directed block this is done by multiple regression, in an outwards directed block by simple regression.

Table 2 Weighting schemes for the inside approximation

| Weighting Scheme no. | Model of the inner relations | $q_{kk'}$ = o, but if $\eta_k$ is | |
|---|---|---|---|
| | | source of $\eta_{k'}$, $q_{kk'}$ = | target of $\eta_{k'}$, $q_{kk'}$ = |
| 1 | Path model, LISREL model | $p_{kk'}$ | $r_{kk'}$ |
| 2 | Path model (Noonan & Wold, 1977) | $p_{kk'}$ | 1 |
| 3 | Path model (Wold, 1977) | $p_{kk'}$ | $p_{k'k}$ sign($r_{kk'}$) |
| 4 | Path model (Wold, note 3) | sign($r_{kk'}$) | sign($r_{kk'}$) |
| 5 | Principal component model of inner relations, MSFA model for total model | $r_{kk'}$ | $r_{kk'}$ |

sign($a$) denotes a 1 with the sign of $a$.

The inside approximation is a weighted combination of the other factors, and it is ambiguous which weights should be applied for

$${}_o\ddot{\eta}_k = \sum_{k'} q_{kk'}\, {}_o\tilde{\eta}_{k'} \quad . \tag{4o}$$

Herman Wold has made several proposals; some of them are listed in table 2. From our considerations above it follows that the factors which are sources of $\eta_k$ should be weighted with the regression coefficient $p_{kk'}$, and the factors which are targets of $\eta_k$ should be weighted by their correlation coefficients (see weighting scheme no. 1 in table 2). Thus the inner approximation ${}_o\ddot{\eta}_k$ tends to become the best mediating variable between the source variables and the target variables. Using the correlation for both the source and target variables (see weighting scheme no. 5 in table 2) serves for maximizing the variance of the principal component of all $\eta_k$ ($k = 1,\ldots,K$), and thus supplies us with a multi-set factor solution.

## 4.4 Orthogonality of factors

In order to obtain the orthogonality between specified pairs of factors, transformations must be performed in each iteration cycle. The PLS algorithm has three points, where these transformations may be inserted (cf. appendix 1):

- as step (6): The outside approximation $o\tilde{\eta}_k$ is orthogonalized on nonrelated factors.
- as step (8b) after step (8): The inside approximation $o\ddot{\eta}_k$ is orthogonalized on the nonrelated factors.
- as step (1o) before ending an iteration cycle: All new approximations are orthogonalized on their nonrelated factors.

It has been found that step (6) has the best effect and step (8b) the poorest. Step (1o) must also be carried out, or else an exchange of the blocks would strongly influence the results.

The orthogonalization is carried out by a transformation matrix $\underset{\sim}{T}$.

$$o\eta^{*} = o\underset{\sim}{T}' \; o\eta \qquad (41)$$

It is stated in appendix 2, how to obtain a transformation matrix $\underset{\sim}{T}$ which provides a factor correlation matrix with zero coefficients in positions specified by the inner-model matrix $\underset{\sim}{M}_I$. The transformation procedure implies an iteration which is stopped when an accuracy criterion, say $\delta_T$, is reached. Thus an iteration procedure is performed within an iteration cycle. It seems to make no sense to require an accuracy in the inner iteration which has not yet been reached in the outer iteration procedure. Therefore, the accuracy reached in the PLS procedure during iteration cycle $o$, say $\delta_o$, is imposed on the transformation procedure. In this way the accuracy of the transformation increases as the accuracy in the PLS procedure increses. This method has been found to work well.

Orthogonality of factors of different blocks of variables is always possible in a two-block model, e.g. in the canonical correlation model. But it must be mentioned that it is not al-

ways possible to compel zero correlations between factors of different blocks. The problem is known from multi-set factor analysis (Horst, 1965). Correlations of factors from different sets can be forced to be low, but not to be zero. However, in the final step of the computations it is possible to abandon the block model, i.e. to ignore the outer-model matrix, and to determine factor having a factor correlation matrix exactly consistent with the inner-model matrix. These factors will have small, but non-zero weights in all indicator variables (cf section 5.3).

# 5. Applications

## 5.1 A path model: Noonan & Wold (1977)

Table 1 shows the parameter estimates for three different three-block models for the problem, shown in figure 1 (Noonan & Wold, 1977). The first model (column 4 of table 1) has only inwards directed measurement relationships. The parameters are estimated with weighting scheme no.2 (cf. table 2). The results are identical to those reported by Noonan and Wold (1977), but in table 1 the factor loadings are reported instead of the factor weights.

The first model (column 4 of table 1) has only inwards directed measurement relationships. Using weighting scheme no. 5, the parameters change only a little, but are parameters of the multi-set factor analysis. The second model (column 5 of table 1) is the model as shown in figure 1. The third model has only outwards directed blocks. The main result is shown in the last two rows. The proportion of variance explained in the inner model ($r_3^2$) and in the outer model (last row) depends on the measurement directions. If the measurement models are inwards directed, only inner residuals are to be minimized, and the squared multiple correlation is high. If the measurement models are outwards directed, the outer residuals are minimized, and the proportion of variance explained in the outer variables is high. This result was also observed in two-block models.

Table 3 Multi-set factor analysis of three test batteries.
Data from Thurstone & Thurstone (1941), cited from Horst (1965).
Block model $M_B$ = (1,1,1).

| Factors/variables | model | within-block factors | across-block factors |
|---|---|---|---|
| | inner model $M_I$ | factor correlations (×1000) | factor correlations (×1000) |
| $\eta_{11}$ | 1 1 1 o o o | 1000 sym. | 1000 sym. |
| $\eta_{12}$ | 1 1 1 o o o | 735 1000 | 737 1000 |
| $\eta_{13}$ | 1 1 1 o o o | 756 743 1000 | 758 745 1000 |
| $\eta_{21}$ | o o o 1 1 1 | o 22 -29 1000 | o o o 1000 |
| $\eta_{22}$ | o o o 1 1 1 | 33 o -33 604 1000 | o o o 604 1000 |
| $\eta_{23}$ | o o o 1 1 1 | 27 -19 o 502 636 1000 | o o o 503 636 1000 |
| | outer model $M_O$ | factor weights (×1000) | factors weights (×1000) |
| battery 1 | | | |
| $Y_{11}$ - verbal | 1 o o 1 o o | 509 o o -921 o o | 494 28 -15 -914 -15 -5 |
| $Y_{12}$ - numerical | 1 o o 1 o o | 355 o o 430 o o | 362 -13 7 437 -10 -5 |
| $Y_{13}$ - spatial | 1 o o 1 o o | 498 o o 491 o o | 506 -15 8 500 -14 -7 |
| battery 2 | | | |
| $Y_{21}$ - verbal | o 1 o o 1 o | o 551 o o -846 o | 24 552 -24 -16 -847 16 |
| $Y_{22}$ - numerical | o 1 o o 1 o | o 499 o o 460 o | -13 499 13 -16 460 15 |
| $Y_{23}$ - spatial | o 1 o o 1 o | o 442 o o 368 o | -11 443 10 -14 369 13 |
| battery 3 | | | |
| $Y_{31}$ - verbal | o o 1 o o 1 | o o 437 o o -958 | 12 -28 457 7 13 -968 |
| $Y_{32}$ - numerical | o o 1 o o 1 | o o 505 o o 433 | -5 12 498 8 14 424 |
| $Y_{33}$ - spatial | o o 1 o o 1 | o o 404 o o 420 | -5 12 396 6 11 413 |

## 5.2 A multi-set factor analysis.

Table 3 shows the results of a multi-set factor analysis, computed by a PLS-program. The correlation matrix is taken from Thurstone and Thurstone, and is used by Horst (1965) for demonstrating his four methods of multi-set factor analysis. The variables are nine subtests, grouped in three test batteries. The inner-model matrix $\underset{\sim}{M}_I$ represents the assumption that the correlations between the factors are strictly canonical, i.e. that nonanalogous factors of different blocks are uncorrelated.

The results show that this assumptions is not supported by the data. If the outer-model is observed exactly and the factors are determined within the blocks, nonanalogous factors of different sets show small correlations. But if the factors are determined across the blocks, the inner model can be observed axactly.

Horst gave the results of the four methods of MSFA (1965, p. 571-584). There are small differences between the results of the four methods. Our results (table 3) lie within the range of the differences between the Horst methods.

Table 4 Correlation matrix for the model on social class, environment and mental abilities. Data from Marjoribanks (1972) and Brandstätter (1976).

| Block | variable | correlation (×100) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Soc.ec. background | Education of father | 100 | | | | | | | | | | | | | | | | | | |
| | Education of mother | 63 | 100 | | | | | | | | | | | | | | | | | |
| | Occupation of father | 79 | 60 | 100 | | | | | | | | | | | | | | | | |
| family structure | Number of children | -9 | -25 | -13 | 100 | | | | | | | | | | | | | | | |
| | Sibling position | -11 | -26 | -12 | 87 | 100 | | | | | | | | | | | | | | |
| | Crowding ratio | -17 | -34 | -25 | 83 | 73 | 100 | | | | | | | | | | | | | |
| Parental dominance | Father dominance | 26 | 30 | 20 | -19 | -22 | -22 | 100 | | | | | | | | | | | | |
| | Mother dominance | 23 | 21 | 20 | -1 | 4 | 1 | -48 | 100 | | | | | | | | | | | |
| Environ-mental forces, press for: | achievement | 50 | 52 | 56 | -46 | -40 | -51 | 32 | 26 | 100 | | | | | | | | | | |
| | activities | 48 | 49 | 45 | -30 | -26 | -33 | 43 | 30 | 72 | 100 | | | | | | | | | |
| | intellectuality | 50 | 54 | 51 | -35 | -31 | -40 | 38 | 37 | 80 | 80 | 100 | | | | | | | | |
| | independence | 37 | 32 | 36 | -6 | -8 | -7 | 11 | 35 | 40 | 46 | 48 | 100 | | | | | | | |
| | English | 32 | 43 | 34 | -16 | -15 | -15 | 23 | 32 | 52 | 63 | 61 | 49 | 100 | | | | | | |
| | ethlanguage | 44 | 46 | 36 | -22 | -22 | -31 | 34 | 21 | 51 | 64 | 62 | 41 | 44 | 100 | | | | | |
| Mental abilities | Verbal | 29 | 38 | 43 | -32 | -27 | -34 | 16 | 21 | 66 | 52 | 61 | 42 | 50 | 35 | 100 | | | | |
| | Number | 26 | 32 | 30 | -33 | -26 | -34 | 10 | 15 | 66 | 41 | 53 | 34 | 27 | 24 | 66 | 100 | | | |
| | Spatial | 26 | 20 | 31 | -4 | -4 | -7 | 9 | 4 | 28 | 22 | 26 | 10 | 18 | 9 | 34 | 38 | 100 | | |
| | Reasoning | 22 | 15 | 29 | -3 | -4 | -9 | 11 | 4 | 39 | 26 | 30 | 22 | 28 | 18 | 39 | 44 | 43 | 100 | |

## 5.3 Social class, environment and mental abilities.

Table 4 shows the correlation matrix of 18 variables, grouped in five blocks of variables. The correlation matrix has been published by Brandstätter (1976), the data gathered by Marjoribanks (1972). The sample consists of 185 11-year-old boys and their parents. Boys where tested by the primary mental abilities test (block 5). Their parents where interviewed in an semistructured interview concerning social status characteristics (block 1), familiy structure (block 2), and learning environment of the home (block 3 and 4). The environmental variables are divided in two blocks, parental dominance (i. e. father's / mother's involvement in child's activities, father's / mother's role in family decision making), and social press. The variables "press for achievement", "press for activities", "press for intellectuality" etc. are assumed to measure two uncorrelated factors.

Table 5 Factor weights and factor structure of the model on social class, environment and mental abilities. Data from Marjoribanks (1972) and Brandstätter (1976).

| Variable | Weights (×100) | | | | | | Correlations (× 100) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\eta_1$ | $\eta_2$ | $\eta_3$ | $\eta_4$ | $\eta_5$ | $\eta_6$ | $\eta_1$ | $\eta_2$ | $\eta_3$ | $\eta_4$ | $\eta_5$ | $\eta_6$ |
| Education of father | 11 | 0 | 0 | 0 | 0 | 0 | 82 | 17 | 48 | 43 | 34 | 33 |
| Education of mother | 60 | 0 | 0 | 0 | 0 | 0 | 92 | 34 | 50 | 45 | 35 | 37 |
| Occupation of father | 41 | 0 | 0 | 0 | 0 | 0 | 86 | 25 | 39 | 51 | 27 | 43 |
| Number of children | 0 | -3 | 0 | 0 | 0 | 0 | -21 | -85 | -20 | -46 | -6 | -29 |
| Sibling position | 0 | -4 | 0 | 0 | 0 | 0 | -22 | -76 | -18 | -40 | -6 | -24 |
| Crowding ratio | 0 | -95 | 0 | 0 | 0 | 0 | -33 | -100 | -21 | -51 | -6 | -32 |
| Father dominance | 0 | 0 | 99 | 0 | 0 | 0 | 29 | 22 | 53 | 23 | 32 | 15 |
| Mother dominance | 0 | 0 | 97 | 0 | 0 | 0 | 23 | -1 | 49 | 18 | 35 | 17 |
| Press for achievement | 0 | 0 | 0 | 130 | -96 | 0 | 61 | 51 | 57 | 95 | 29 | 70 |
| Press for activities | 0 | 0 | 0 | -39 | 70 | 0 | 53 | 33 | 72 | 48 | 77 | 49 |
| Press f. intellectuality | 0 | 0 | 0 | -8 | 76 | 0 | 59 | 40 | 74 | 65 | 71 | 60 |
| Press f. independence | 0 | 0 | 0 | -2 | 22 | 0 | 38 | 7 | 45 | 28 | 57 | 39 |
| Press for English | 0 | 0 | 0 | -6 | 16 | 0 | 43 | 15 | 54 | 34 | 64 | 43 |
| Press for Ethlanguage | 0 | 0 | 0 | 6 | -7 | 0 | 47 | 31 | 54 | 40 | 52 | 31 |
| Verbal ability | 0 | 0 | 0 | 0 | 0 | 42 | 44 | 34 | 36 | 59 | 34 | 85 |
| Number ability | 0 | 0 | 0 | 0 | 0 | 41 | 35 | 34 | 24 | 65 | 16 | 87 |
| Spatial ability | 0 | 0 | 0 | 0 | 0 | 18 | 28 | 7 | 13 | 25 | 13 | 59 |
| Reasoning ability | 0 | 0 | 0 | 0 | 0 | 25 | 23 | 9 | 15 | 37 | 12 | 68 |

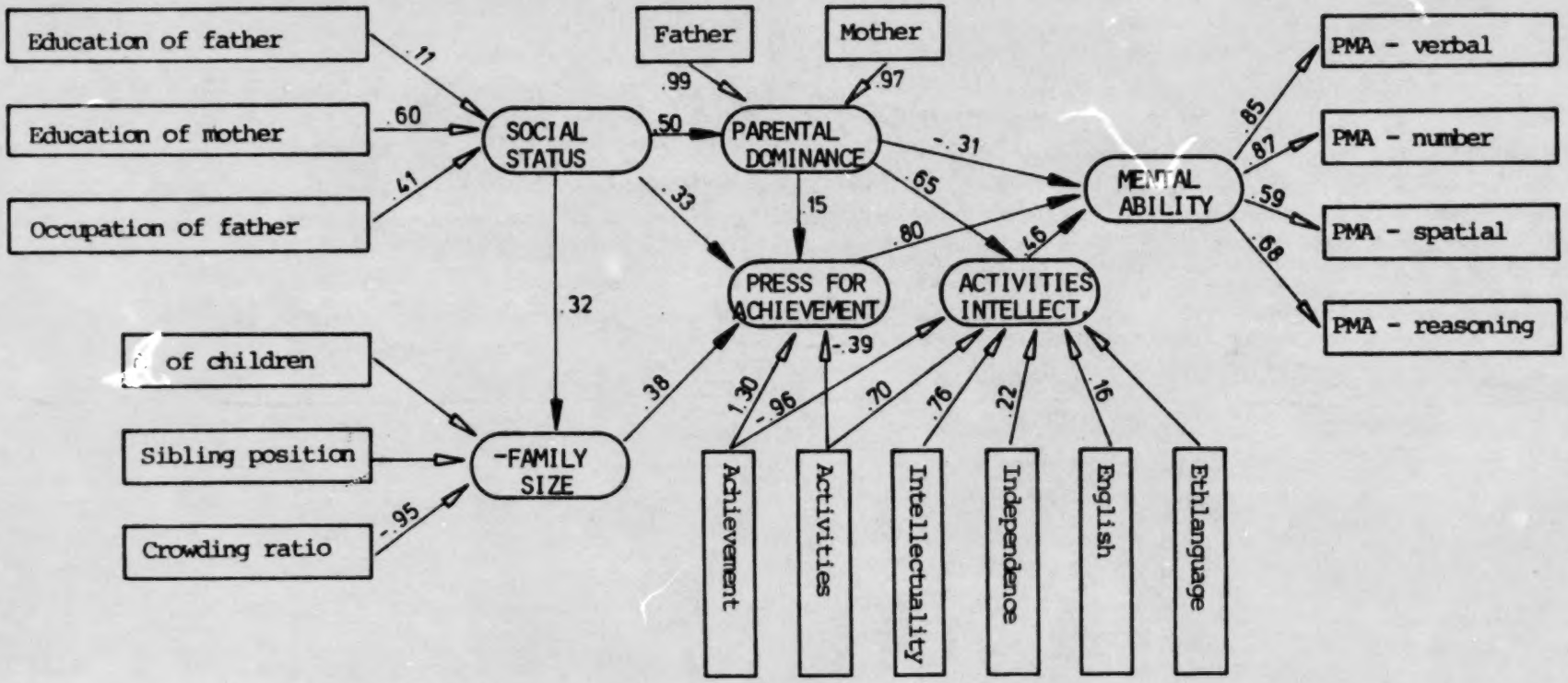


Figure 5: Path diagram of the model on social class, environment, and mental abilities. Path coefficients and weights less than |.10| have been omitted.

Table 5 shows the factor weights and the factor structure. The weights of the variables education of father, number of children, sibling position, and press for ethlanguage (refering to any language spoken at home other than English) are very low, and these variables could have been omitted. Table 6 shows the correlations and the path coefficients of the factors.

Table 6 Factor correlations and path coefficients of the model on social class, environment and mental abilities. Data from Marjoribanks (1972) and Brandstätter (1976).

| Structural variable | Correlation (×100) | | | | | | Path coefficient (×100) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\eta_1$ | $\eta_2$ | $\eta_3$ | $\eta_4$ | $\eta_5$ | $\eta_6$ | $\eta_1$ | $\eta_2$ | $\eta_3$ | $\eta_4$ | $\eta_5$ | $\eta_6$ |
| $\eta_1$ - Social status | 100 | | | | | | - | - | - | - | - | - |
| $\eta_2$ - Family size | 32 | 100 | | | | | 32 | - | - | - | - | - |
| $\eta_3$ - Parental dominance | 52 | 21 | 100 | | | | 50 | 5 | - | - | - | - |
| $\eta_4$ - "Press f.achievem" | 53 | 51 | 40 | 100 | | | 33 | 38 | 15 | - | - | - |
| $\eta_5$ - "Press f.activities" | 36 | 6 | 66 | 01 | 100 | | 5 | -9 | 65 | - | - | - |
| $\eta_6$ - Mental ability | 44 | 32 | 31 | 66 | 26 | 100 | 3 | -7 | -31 | 80 | 46 | - |
| Squared multiple correlation ($R^2$×100) | | | | | | | 0 | 11 | 27 | 43 | 44 | 54 |

The variables father's dominance and mother's dominance, though negatively correlated, contribute with positive weights to the factor parental dominance. The explanation of this fact can be found in the correlation matrix (table 4). Both variables correlate not with different, but with equal sign with most of the other variables. Therefore a factor "parental dominance" fits better into the path model than a factor "mother's dominance versus father's dominance". The direct influence of parental dominance on mental abilities is negative ($p_{63} = -.31$), whereas the total influence is positive ($r_{63} = .31$). That is to say that the pure influence of dominance hampers the development of mental abilities, whereas dominance stimulates mental abilities when related to substantial goals of education.

The two social-press factors show really different correlations and path coefficients to the other factors. The factor "press for achievement" is dependent on social status and family size and highly predictive for mental abilities, whereas the

factor "press for activity and intellectuality" is dependent on parental dominance and not as predictive for mental abilities.

Main result is that social class -- though fairly correlated -- has no direct influence on mental abilities, if the learning environment is taken into consideration.

## 5. Discussion

5.1 Wold's PLS method has been extended so as to cope with (a) fixed zero-correlations between specified pairs of factors, (b) simultaneous extraction of several factors from the same block of indicator variables, (c) different numbers of factors from different blocks, and (d) multi-set factor models.

5.2 The consistency of PLS estimates is discussed by Wold (note 3) and proved by simulation studies (Hui, 1978; note 1). "Convergence of NIPALS estimation has never been a problem" (Wold, 1977, p.546). This applies also for the extensions presented in this paper.

5.3 Differences and relative merits of the PLS and the ML approach to linear structural relation models are discussed by Wold (1975, p. 349; note 2).

5.4 It is not recommended to use the extended PLS algorithm for principal component analysis, canonical correlation analysis, and multi-set factor analysis. Convergence may be very slow if many factor correlations are fixed to zero. But the results are identical to those of the Wold PLS algorithm, using the deflation technique and successive estimation of dimensions, and to those of standard programs. This shows that the method works correctly.

5.5 The algorithm proposed here can be simplified with respect to the treatment of inwards and outwards directed blocks. Let $\underset{\sim}{S}^{+} = \underset{\sim}{T}' \underset{\sim}{S} \underset{\sim}{T}$ the covariance matrix of indicator variables, all variables in inwards blocks being transformed to orthogonality. Then $\underset{\sim}{T}$ must be a supermatrix $(L \times L)$ with elements $\underset{\sim}{t}_{ll'}$ $(J_l \times J_{l'})$

$$\underset{\sim}{t}_{ll'} = \begin{cases} \underset{\sim}{0} & \text{if } l \neq l' \\ \underset{\sim}{I} & \text{if } \underset{\sim}{Y}_l \text{ is outwards directed} \\ E^{-1/2}(\underset{\sim}{Y}_l \underset{\sim}{Y}_l') & \text{if } \underset{\sim}{Y}_l \text{ is inwards directed} \end{cases}$$

(cf. Horst, 1975, p. 569). Using $S^{+}$ for the parameter estimation, all blocks can be considered as outwards directed blocks. In a final step the factor weights and factor structure of the transformed variables must be related to the untransformed variables by $B = T^{-1} B^{+}$ and $W = T^{-1} W^{+}$.

5.6 The PLS method may be extended so as to deal with fixed elements in the factor weights or the factor pattern matrix. This can not be done, if the simplification of §5.5 is implemented.

## Appendix 1: The PLS algorithm for several factors measured by the same block of indicator variables.

Definition of notation

| | | |
|---|---|---|
| I | | number of observations, $i,i' = 1,\ldots,I$ |
| J | | number of variables, $j, j' = 1,\ldots,J$ |
| K | | number of factors, $k,k' = 1,\ldots,K$ |
| L | | number of blocks of variables, $l, l' = 1,\ldots,L$ |
| Y | (J×I) | data matrix; $\sum_i y_{ij} = o$ |
| $\hat{\eta}$ | (K×I) | factor score matrix |
| $Y_k$ | ($J_k$×I) | part of data matrix, measuring factor $\eta_k$ |
| $Y_l$ | ($J_l$×I) | part of data matrix, containing variables of block l |
| $M_B$ | (L) | Block-model matrix |
| $M_I$ | (K×K) | Inner-model matrix |
| $M_O$ | (J×K) | Outer-model matrix |
| S | (J×J) | $= 1/I\ (Y\ Y')$ covariance matrix of variables |
| R | (K×K) | $= 1/I\ (\hat{\eta}\ \hat{\eta}')$ correlation matrix of factors |
| B | (J×K) | $= 1/I\ (Y\ \hat{\eta}')$ covariance of variables and factors, factor structur |
| W | (J×K) | $\rightarrow \hat{\eta} = W'\ Y$ factor weights in variables |
| P | (K×K) | $\rightarrow \hat{\eta} = P\hat{\eta} + \hat{\zeta}$ path coefficient matrix, where $\hat{\zeta}$ are residuals |
| $S_k^{-1}$ | (J×J) | Inverse of $1/I(Y_k\ Y_k')$ (K×K), but extended to order (J×J) by appropriatly inserted zero vectors. |
| T | (K×K) | Transformation matrix. The expression "T = transform($M_I$,R,κ)" refers to a procedure which determines a transformation matrix T so that T' R T has zero elements in the same position as $M_I$. If κ = k, the transformation is determined only for the k-th factor. If κ = o, the transformation is determined for the total matrix. |
| Q | (K×K) | inner weight matrix. The expression "Q=f($M_I$,R,P,weighting scheme)" refers to the different ways to define the inner weight matrix Q. see table 2. |
| P | (K×K) | path coefficient matrix. The expression "P=pathanalysis($M_I$,R)" refers to a procedure which delivers path coefficients P = P * $M_I$ for a correlations matrix R. |
| $\eta_k$, $T_k$, $W_k$ | | The k-th column of $\eta$, T, W, resp. |

The computing formulas are written out as equations on the data matrix Y and the factor score matrix $\hat{\eta}$ on the left hand side, and as assigment statements, related to the covariance matrices S, R und B, on the right hand side. Here the pre-subscript c is omitted from the matrices W, R, B, T and Q, as well as the different diacritics (asterisk, tilde, trema).

## The PLS algorithm

Step

(1) Prerequisites are the model matrices $M_B$, $M_I$, $M_O$, the data matrix Y or the covariance matrix S, and an initial approximation to the factor weights, e.g. ${}_oW = M_O$ .

(2) Do step (3) to step (11) for each iteration cycle c, c = 1,...

(3) Do step (4) to step (9) for each factor $\eta_k$, k = 1,...,K

(4) Define factor $\eta_k$ in terms of its indicator variables (outside approx.)

$$ {}_c\tilde{\eta}_k = {}_{c-1}W_k \; Y_k $$

W := W diag$^{-1/2}$(W' S W)

R := W' S W

B := S W

(5) If convergence has occured, go to step (12).

(6) Orthogonalize $\tilde{\eta}_k$ on nonrelelated factors.

$$ {}_c\tilde{\eta}^{\bullet}_k = {}_cT'_k \; {}_c\tilde{\eta}_k $$

T := transform($M_I$,R,k)

R := T' R T

B := B T

(7) Define factor $\eta_k$ in terms of other factors (inside approximation).

$$ {}_c\ddot{\eta}_k = {}_cQ'_k \; {}_c\tilde{\eta}^{\bullet}_k $$

P := pathanalysis($M_I$,R,k)

Q := f($M_I$, R, P, weighting scheme)

R := T' R T

B := B T

(8) Compute weights by regression

B := B * $M_O$

in an outwards directed block:

$$ {}_cW_k = 1/I \; (Y_k \; {}_c\ddot{\eta}'_k) $$

$W_k := B_k$

in an inwards directed block:

$$ {}_cW_k = (Y_k \; Y'_k)^{-1} \; (Y_k \; {}_c\ddot{\eta}'_k) $$

$W_k := S_k^{-1} \; B_k$

(9) End of computation of factor $\eta_k$ .

(10) Orthogonalize all factors on all nonrelated factors.

$$ {}_c\hat{\eta} = {}_cT' \; {}_c\tilde{\eta} $$

T := transform($M_I$, R, o)

R := T' R T

B := B T

W := W T

(11) End of iteration cycle c .

(12) Final computations

P := pathanalysis ($M_I$, R, o)

W := $S^{-1}$ B

(13) Stop

## Appendix 2: Transformation to obtain specified factor correlations equal to zero.

To obtain a factor correlation matrix R with an specified pattern of zeros, an algorithm is proposed resembling the Jacobi method for diagonalizing a symmetric matrix (cf. Bock, 1975, p. 96).

Let R the covariance matrix of a set of variables $X_1,\ldots,X_k,X_l,X_m,\ldots,X_M$. Then an orthonormal transformation $T_{kl}$ (2×2) may be applied on the variables $X_k$ and $X_l$ to abtain uncorrelated substitutes for them. Each pair of variables is orthogonalized by a transformation. But performing the orthogonalization on a pair $X_k$ and $X_l$ may destroy the zero-correlation between a pair $X_k$ and $X_m$. So if the transformations on all different pairs are completed, a new cycle of transformation must be begun, starting anew with the first pair of variables. After some cycles the process converges, and the covariance between each pair of variables differs from zero not more than a specified small value delta. Then the diagonal contains the eigenvalues of the covariance matrix.

The algorithm to be proposed here differs in two respects from the Jacobi method:

- The transformation is applied only on those pairs of variables whose correlations are restricted by the model to be zero.
- The method is applied to a correlation matrix, and nonsingular transformations are used to preserve the unitiy diagonal of the correlation matrix. From the infinite number of transformations $T_{kl}$ fulfilling this requirement we consider two special cases, the semipartial and the symmetric transformation.

The semipartial transformation ${}_pT_{kl}$ partials out $X_l$ from $X_k$ and retains $X_l$ unaltered,

$$ {}_pT_{kl} = \begin{pmatrix} 1/\sqrt{1-r_{kl}^2} & 0 \\ -r_{kl}/\sqrt{1-r_{kl}^2} & 1 \end{pmatrix}, $$

where $r_{kl}$ is the correlation of $X_k$ and $X_l$.

The symmetric transformation ${}_sT_{kl}$ transforms both $X_k$ and $X_l$ so as the transformed variables to be uncorrelated but least-squares approximations to their originals (see Kaiser, 1967).

$$ {}_sT_{kl} = \begin{pmatrix} P+M & P-M \\ P-M & M+P \end{pmatrix} $$

where $P = 0.5/\sqrt{1+r_{kl}}$ and $M = .5 / \sqrt{1-r_{kl}}$ .

The transformation

$$ (X_k^*, X_l^*) = (X_k, X_l)\, T_{kl} $$

bothers the correlations of $X_k$ and $X_l$ with another variable $X_m$ .

$$ (r_{km}^*, r_{lm}^*) = (r_{km}, r_{lm})\, T_{kl} $$

All transformation in pairs may be collected in one transformation matrix for all variables, say $T$ $(M \times M)$, where $M$ is the total number of variables. Let $T_{jk}^i$ a unity matrix of order $(M \times M)$, but the elements with the indices kk, kl, lk, ll replaced by the elements of the $2 \times 2$ transformation matrix $T_{jk}$, as applied on the variables $X_j$ and $X_k$ in iteration cycle $i$ . Then the total transformation matrix is given by the product of all transformations.

$$ T = \prod_i \prod_{k<l} \prod T_{kl}^i $$

This transformation algorithm has to fulfill fewer restrictions than the Jacobi algorithm. If it is possible to diagonalize a symmetric matrix, then it also is possible to retain some nondiagonal elements of this matrix not necessarily equal to zero. Therefore this algorithm will converge.

Reference Notes

1. Consitency at large in PLS. First investigations by B. Hui; H. Wold & A. Edman; H. Apel & H.Wold. Compiled by H. Apel. Frankfurt, October 1978.
2. Wold, H. Path models with latent variables: Comparative aspects of the PLS (Partial Least Squares) versus the ML (Maximum Likelihood) approach. Paper read at the 1978 European meeting on Psychometrics and Mathematical Psychology, June 15-17, 1978, Uppsala.
3. Wold, H. Model construction and evaluation when theoretical knowledge is scarce: An example of the use of Partial Least Squares. Cahier 79.o6 du département d'économétrie, faculté des sciences économiques et sociales, université de Genève, 1979.

References

Apel, H. Simulation sozio-ökonomischer Zusammenhänge: Kritik und Modifikation von Systems Analysis. Doctoral dissertation, J. W. von Goethe Universität, Frankfurt/Main. Darmstadt: Toeche Mittler, 1976.

Apel, H. Soft models of Herman Wold's type: Models of environmental problems and underdevelopment. To appear in W. Buhr & P. Friedrichs (Eds.), Regional development under stagnation. London. (Paper, Frankfurt, October 1978).

Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Brandstätter, J. Soziale Schicht, Umwelt und Intelligenz: Eine Pfadanalyse der Korrelationsbefunde von Marjoribanks. Psychologische Beiträge, 1976, 18, 35-53.

Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart and Winston, 1965.

Hotelling, H. The most predictable criterion. Journal of Educational Psychology, 1935, 26, 139-142.

Hui, B. S. Soft modelling by partial least squares approach. Doctoral dissertation, University of Pennsylvania, 1978.

Jöreskog, K. G. Structural equation models in the social sciences: Specifications, estimation and testing. In P. R. Krishnaiah (Ed.), Application of statistics. Amsterdam: North Holland, 1977.

Jöreskog, K. G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-477.

Jöreskog, K.G., & Goldberger, A. S. Estimation of models with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association, 1975, 1o, 631-639.

Kaiser, H. F. Uncorrelated linear composites maximally related to a complex of correlated observables. Educational and and Psychological Measurement, 1967, 27, 3-6.

Kettenring, J. R. Canonical analysis of several sets of variables. Biometrika, 1971, 58, 433-451.

Marjoribanks, K. Environment, social class, and mental abilities. Journal of Educational Psychology, 1972, 63, 1o3-1o9.

Noonan, R., & Wold, H. NIPALS path modelling with latent variables: Analysing school survey data using nonlinear iterative partial least squares. Scandinavian Journal of Educational Research, 1977, 21, 33-61.

Skinner, H. Exploring relationships among multiple data sets. Multivariate Behavioral Research, 1977, 12, 199-22o.

Tucker, L. R. An inter-battery method of factor analysis. Psychometrika, 1958, 23,111-136.

van den Wollenberg, A. L. Canonical redundancy analysis: An alternative for canonical correlation analysis. Psychometrika, 1977, 42, 2o7-219.

von Mises, R., & Pollaczek-Geiringer, H. Praktische Verfahren der Gleichungsauflösung. Zeitschrift für angewandte Mathematik und Mechanik, 1929, 9, 58-77 and 152-164.

Wold, H. Nonlinear estimation by iterative least squares procedure. In F. N. David (Ed.), Research papers in statistics: Festschrift for J. Neyman. New York: Wiley, 1966. (a)

Wold, H. Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiah (ed.), Multivariate Analysis. New York: Academic Press, 1966. (b)

Wold, H. Nonlinear Iterative Partial Least Squares (NIPALS) modelling: Some current developments. In P. R. Krishnaiah (Ed.), Multivariate analysis-III. New York: Academic Press, 1973.

Wold, H. Path models with latent variables. In H. M. Blalock & al. (Eds.), Quantitative sociology. New York: Academic Press, 1975.

Wold, H. On the transition from pattern recognition to model building. In R. Henn and O. Moeschlin (eds.), Mathematical economics and game theory: Essays in honor of Oskar Morgenstern. Berlin, Heidelberg, New York: Springer, 1977.

Wold, H. Open path models with latent variables. In H. Albach, E. Helmstädter, R. Henn (Eds.), Quantitative Wirtschaftsforschung: Wilhelm Krelle zum 6o. Geburtstag. Tübingen: Mohr, 1977.